# 中国计算档案学发展的SWOT分析与策略研究*

■ 赵跃[1] 马晓玥[1] 张佳欣[2]
[1] 四川大学公共管理学院 成都 610065 [2] 中国人民大学信息资源管理学院 北京 100872

摘 要：[目的/意义]回顾计算档案学发展现状，探索计算档案学在中国的发展策略，为新文科建设背景下中国计算档案学的发展提供参考。[方法/过程]基于文献研究梳理计算档案学的发展现状，运用SWOT分析法剖析计算档案学在国内发展的机会、威胁、优势、劣势等内外部环境要素，并通过内外要素交叉匹配形成不同发展策略。[结果/结论]研究发现，计算档案学的发展在国际上得到持续关注，且围绕基础理论、教育研究、档案处理、档案分析、档案化处理等方面初步确立了理论研究与实践探索的方向，但总体上，计算档案学的发展尚处于初步探索阶段。中国推进计算档案学需要注重：依托交叉学科的建设机遇，打造超学科研究平台；瞄准领域信息化战略需求，形成规模化研究方向；厘清与相关学科的边界，突出计算档案学的特色；发挥超学科的研究优势，规避数据安全隐私风险；抓住复合型人才培养契机，整合多方资源共建教学平台；围绕实践领域的核心问题，探索可操作的技术解决方案；加强制度设计与技术攻坚，做好档案安全风险评估与管控；加大基础研究，明晰计算档案学的理论、方法和技术体系。

关键词：计算档案学 新文科 数据转型 档案教育
分类号：G270
DOI：10.13266/j.issn.0252-3116.2022.04.006



## 1 引言

自2006年以来，云计算、大数据、人工智能等技术成为推动社会演进的主要动力。这些新兴技术产业提供的技术更新、思想支撑和研究方法，使得计算思维的运用跳出经济的范畴，衍生出更多与之相关的思维模式和跨领域实践，计算社会科学、计算语言学等“计算+学科”成为大数据时代的新兴学科范式。随着数智时代的到来，日趋复杂的技术环境同样也让人们认识到传统的文件与档案管理实践要实现数字转型就需要计算理论方法的引入；而新的档案形式和档案问题的出现，不仅需要计算机科学等学科的介入，更需要档案理论方法的助力。在这种新的双向需求刺激下，强调计算理论方法与档案理论方法融合的计算档案学兴起具有必然性。

一方面，新兴的原生数字档案的生产和消费由社会和工业趋势以及与档案方法几乎没有联系的计算机和数据方法所决定。要了解它们的生产和消费特征、治理要点，解决规模化原生数字档案的处理、分析、存储、长期保存和获取问题，就需要计算机科学等学科的助力，尤其是计算方法和资源的应用。同时，要确保新形式文件和档案的真实性、完整性、可靠性、可用性、安全性等特征，也需要档案学的介入。面对新的档案形式，多学科领域共同应对是必然趋势。另一方面，传统档案工作正加速推进数智转型，档案数据驱动的研究活动兴起，使得大规模档案材料的数据化加工、处理、关联、分析、挖掘等成为挑战。同时，技术赋能档案工作的需求愈发凸显，文件归档、开放鉴定和敏感性审查等工作实现自动化、智能化以提升工作效率的目标逐渐清晰。总之，面对新的档案实践，传统档案领域面临理论、方法、技术的局限，传统档案理论、方法和技术无法解决新的档案实践带来的大规模处理与应用的挑战，当代档案工作需要计算思维和方法的引入。

为深入探索计算与档案在思维、理论、方法中的高级融合形式，在2016年美国马里兰大学举办的以“发现新知识：大数据时代的档案文件”为主题的计算档案学专题研讨会上，计算档案学（Computational Archival Science）概念被提出，随后召开的首届IEEE计算档案

* 本文系国家社会科学基金项目“面向‘三化融合’的非遗档案资源建设多元协同模式研究”（项目编号：20CTQ034）研究成果之一。
作者简介：赵跃，副研究员，博士，E-mail：zhaoyuexxe@scu.edu.cn；马晓玥，硕士研究生；张佳欣，硕士研究生。
收稿日期：2021-07-11 修回日期：2021-11-25 本文起止页码：56-66 本文责任编辑：王传清

学工作坊将其初步认定为交叉学科,该学科旨在将计算方法和资源应用于大规模文件/档案的处理、分析、存储、长期保存和获取,以提高效率、生产力和精确度,便于支持鉴定、整理和描述、保存和获取决策以及使用档案材料进行研究[1]。随后,国外学者对此概念涉及的学术领域进行了持续、广泛的探索。近年来,国内学者也注意到国外计算档案学的发展趋势,傅天珍于2019 年发文总结了计算档案学的发展历程、定义和特征[2]。随后,周文泓等[3]、陶玉芳[4]、刘越男等[5]、于英香等[6]也基于文献研究方法对国外计算档案学发展情况进行了多角度的透视。赵跃等[7]基于中国图情档界的计算档案学认知调查,剖析了计算档案学在中国的发展前景。本文将在相关研究基础上,借助 SWOT 分析法探讨计算档案学在中国的发展策略,以期对此领域的进一步研究与实践有所启示。

## 2 计算档案学发展现状回顾

### 2.1 作为一个超学科研究领域得到持续关注

刘越男等[5]对计算档案学在国外的发展进程进行了如下系统梳理:2015 年美国马里兰大学信息学院 R. Marciano 教授组建的探讨利用计算方法解决档案问题的小型跨学科研究小组,被认为是国际上计算档案学的起源。2016 年4 月在马里兰大学召开的计算档案学专题研讨会,宣告跨地域跨学科的学术社区初步形成,6 位会议发起人此后一直是此领域的中坚力量,受邀代表来自英国、加拿大、南非和美国的高校、研究机构、政府机构、文化机构和合作组织。经过 4 年的发展,计算档案学社区进一步吸纳了美国多个高校、州档案馆以及巴西利亚大学、英国阿兰·图灵研究院、阿姆斯特丹大学、印度古吉拉特邦中央大学、印度管理研究所、日本九州大学、堪培拉大学等各国高校和研究机构的研究人员。该学术社区通过举办学术会议、发布专刊、开展合作研究等方式不断扩大规模,加深探索,推动计算档案学的发展。在学术会议方面,据不完全统计,2016-2020 年间计算档案学学术社区以工作坊的形式发起过 27 场学术活动,不乏由知名计算科学研究机构主办的学术会议,如英国阿兰·图灵研究院 2020 年举办的计算档案学讨论会。其中最有代表性的是始于2016 年的 IEEE 大数据计算档案学工作坊,至今已连续举办 5 届,体现了以计算科学为主要阵地的大数据领域对跨学科的计算档案学的重视,并吸引了越来越多来自计算机科学、档案学、信息科学、图书馆学、历史学、艺术学等领域的学者加入,计算档案学的学术社区持续扩大。

IEEE 大数据计算档案学工作坊自 2016 年起共产出 62 篇会议论文,发表数量呈年度递增态势,体现出计算档案学受到的持续关注趋势。除此之外,笔者进一步以“computational archival science”为检索词在 Google scholar 以及 Emerald 等学术搜索引擎和数据库进行检索(检索时间为 2021 年 2 月 10 日),经人工判读剔除重复和不相关的记录后,又获得 13 篇有效外文文献。通过对文献作者的统计发现,75 篇文献来自 19 个国家(地区)的 243 名学者,可见计算档案学在发展初期便得到了较多国家研究人员的关注。从作者分布情况来看,19 个国家(地区)中,来自美国的学者最多,有 163 名,其次为加拿大(21 名)和英国(9 名);243 名学者中,来自马里兰大学信息学院的教授发文较多,R. Marciano 发文最多,有 14 篇,其次为 W. Underwood (9 篇)和 G. Jansen(6 篇);发文数量 3 篇及以上的作者均来自图书情报与档案管理领域,可见图书情报与档案管理学科在计算档案学建设中的基础性作用。从作者合作情况来看,75 篇文章中,由 2 名及以上学者合作完成率达到 60%,体现出计算档案学研究具有较高的合作度。从机构合作情况来看,由 2 个及以上机构合作完成的文章数占比为 46.7%,机构合作度较高。在具体的机构合作方式上,10 篇文章由高校间跨校合作完成,4 篇文章由非高校跨机构合作完成,8 篇文章由校内跨院系或单位合作完成,8 篇文章由校内外跨机构合作完成,3 篇文章由跨校跨机构多方合作完成,1 篇文章由跨院系跨机构多方合作完成,1 篇文章由跨院系跨校跨机构多方合作完成,再次突出了合作,尤其是跨学科的合作对于计算档案学研究的重要性。

此外,75 篇文献共涉及 138 所机构,其中美国马里兰大学信息学院发文最多(17 篇),其次为加拿大英属哥伦比亚大学信息学院(8 篇)、英国伦敦国王学院数字人文系(5 篇),发文量超过 2 篇的 11 个机构多为高校研究机构,可见高校研究机构在计算档案学领域发展中起着核心的推动作用。其中,马里兰大学信息学院、加拿大英属哥伦比亚大学信息学院以及英国伦敦国王学院数字人文系更是计算档案学研究的核心机构。

马里兰大学信息学院致力于开发智慧城市技术,创建新兴的存档方法,其立足于 46 个研究资助项目和多个研究中心(如社会数据科学中心、计算语言学和信息处理中心、人机交互实验室、社区信息高级研究中

心、档案未来中心、Trace 研发中心等[8]），开拓了计算档案学、数字人文、计算语言学、人机交互等 14 个研究领域。其中，2015 年 R. Marciano 教授创立的数字管护创新中心[9]，尤其注重探索档案数据和技术的融合形式，开拓新兴档案分析形式，加深历史、社会、科学和文化研究与档案的结合。自成立以来，该中心与多方合作研究了 10 余个跨学科项目（如与马里兰州档案馆合作开展的奴隶制遗产项目[10]、与 NARA 合作开展的二战档案专题项目[11-12]），成为推动计算档案学发展的中坚力量。

英属哥伦比亚大学信息学院建设了具备智能设备的 Kitimat 实验室、露台实验室、格雷格实验室，以用于网站研究或焦点研讨，还设立了专门的技术顾问提供个性化帮助，涵盖面向 SQL 数据库、网站开发、编程、原型设计等多个技术主题。该学院档案学方向的 V. Lemieux、N. Payne 及各自团队在区块链、人工智能等领域进行了创新性探索，是计算档案学领域的核心力量之一。V. Lemieux 团队开发了基于区块链的处置应用程序和“ArchContracure”智能契约[13]，并应用到在土地交易、医疗记录和财务文件管理等领域[14]。N. Payne 博士设计了分类精准度和文件联系度并重的系统[15]，并开发了支持文件自动分类的新型背景信息捕获框架等[16]。

伦敦国王学院数字人文学系致力于研究数字文化和社会以及用于人文社科研究的先进技术方法[17]，设立了数字文化与数字媒介、数字方法与数字设备、数字社区参与平台与渠道 3 个主要方向。为弥补软件工程和技术管理方面的不足，2015 年还建设了拥有软件工程团队的独立数字实验室，立足于不同行业的实践需求，实验室组建了由研究软件分析师、工程师、UI/UX 设计师、项目经理和系统经理组成的团队，承接了 100 余项数字人文项目（如欧洲大屠杀基础设施项目（European Holocaust Research Infrastructure，EHRI）[18]、欧洲大数据和社会挖掘研究基础设施项目[19]）。此外，2019 年，伦敦国王学院数字人文系与马里兰大学信息学院、马里兰州档案馆、英国国家档案馆合作，成立计算档案学国际研究合作网，开展为期一年的合作以进一步推动计算档案学跨学科探索与实践[20]。

## 2.2 理论研究与实践探索的方向已初步确立

2016 年以来，IEEE 大数据计算档案学工作坊形成了较为稳定的讨论主题：①分析在档案材料中的应用，包括文本挖掘、数据挖掘、情感分析、网络分析；②支持档案处理的分析，包括电子发现、个人信息识别、鉴定、整理和描述；③可扩展的档案服务，包括识别、保存、元数据生成、完整性检查、规范化、对账、关联数据、实体提取、匿名化和缩减；④新的档案形式，包括网络、社交媒体、视听档案和区块链；⑤网络基础设施，用于基于档案的研究以及馆藏的开发和托管；⑥大数据和档案理论与实践；⑦数字管护与保存；⑧众包和档案；⑨大数据以及记忆与身份的构建；⑩特定的大数据技术（例如 NoSQL 数据库）及其应用；⑪大档案数据的语料库和参考集合；⑫关联数据和档案；⑬大数据和来源；⑭从档案中构建大数据研究对象；⑮大数据档案中的法律和道德问题。

这些主题初步罗列出计算档案学相关实践探索方向，后来 R. Marciano 等总结了驱动计算档案学研究的 8 个典型实践：①进化原型和计算语言学；②图形分析与数字人文；③计算机检索工具；④数字管护；⑤公众参与（档案）内容；⑥真实性；⑦网络基础设施和文件连续体；⑧空间和时间分析[21]，进一步勾勒出计算档案学的“领地范围”，并促成一些核心研究领域的形成，如档案材料分析、新形式档案开发、档案化处理的拓展服务、大数据和档案的理论与实践等[22]。但是，要在众多计算和档案研究中划清计算档案学边界非常困难，M. Lee 等提出评估计算档案学研究的启发式方法，用以评估研究问题是否属于此领域核心问题，认为“计算思维解决档案问题”不一定属于计算档案学范畴，计算档案学研究应该以档案和计算问题的共同目标为切入点，融合档案和计算的理论形成自己的专业和新的理论[23]，此方法对于判定计算档案学的核心问题有一定启示，但无法用以准确划分计算档案学研究领域和边界。

通过对国内外计算档案学领域文献研究主题的分析，笔者认为当前计算档案学领域已经初步确立包括基础理论、教育研究、档案处理、档案分析和档案化处理 5 个方向。计算档案学基础理论研究致力于阐释计算档案学相关概念、特征、研究框架与学科属性等。例如，在概念方面，2018 年，R. Marciano 等对 2016 年首届 IEEE 计算档案学工作坊上提出的初步定义[1]进行了更新[21]，将交叉学科（interdisciplinary）更改为超学科（transdisciplinary），强调学科知识的融合。后来，也有学者对计算档案学的定义进行了进一步扩充和阐释[22]。但目前，计算档案学的定义仍是不断发展的，当前定义没有充分反映在超学科下基础学科之间的知识交换，仍具有局限；关于学科属性，一般认为计算档案学是计算机科学与档案学的双向互动，是其要素重

组与融合创造的新超学科领域，但也有学者提出要以档案学、信息科学和计算机科学为基础[22]，甚至有学者指出计算档案学并非一个新的科学领域，而只是一个信息技术方法不断扩大的档案学方向[24]。此外，还有学者强调计算档案学的工程属性，进一步提出档案工程的概念，认为计算档案学的价值只有在提供产品和服务时才能实现[25]。

计算档案学教育研究方向主要关注计算思维与档案思维相关教育、培训与课程设置等问题。H. Stančić 等调查发现，在信息通讯技术影响下，2003－2016 年欧洲高校档案学专业课程设置扩充到信息系统、数字保存等方面，他们认为档案工作者还需进一步学习语义网、图数据库、机器学习等技术[26]；马里兰大学信息学院则提出了由 22 种计算实践构成的计算思维，并分为数据实践、建模和仿真实践、计算问题解决实践和系统思维实践四大类[27]，他们将档案课程中不同的知识单元与计算思维对应，并构建出计算思维关联图书情报硕士教育的两种方式：一是创建新课程教授相关知识领域的计算思维；二是将计算思维引入研究生课程的范例、练习和项目中[28－29]。此外，为促进专业培训的深入，马里兰大学信息学院还致力于建设用以展示、共享和教授档案工作者和研究人员实践的计算档案学教育系统网络平台，通过项目简介、课程计划和案例文件实现用于数字档案的计算案例研究和课程的共享，促进教育者和实践者相互学习[30]。

计算档案学档案处理方向主要探讨档案材料的处理问题，如数字化，电子发现，信息识别、鉴定、分类、整理、描述和访问，数字管护，语义本体，关联数据，主题建模，自然语言处理，机器学习等。例如，在数字化方面，欧洲数据基础设施(European Data Infrastructure，EUDAT)使用 OCR 技术将植物标本图片数字化，采取集成计算分析后转移到可信数字存储库，实现研究数据的共享和长期保存[31]。在分类方面，N. Payne 比较了数字档案自动化分类的方法，提出设计分类精准度和文件联系度并重的系统[15]，还提出以机器学习方法融合司法、历史、程序、业务、技术等不同要素的元数据框架来实现文件的自动分类[16]。在鉴定方面，密歇根大学图书馆通过创建评估选项卡工具对大规模数字档案开展敏感信息的自动识别和鉴定[32]。T. Hutchinson 提出利用自然语言处理技术开展主题建模，帮助识别审查文档的隐私信息[33]或以监督型机器识别个人信息的隐私数据管控隐私风险[34]。在描述组织方面，EHRI 通过收集可信可用的分级档案元数据整合大量分散的大屠杀相关材料数据，并创建 API 目录实现元数据收取、关联、检索功能，在门户网站提供元数据跨国访问[18，35]。马里兰大学以非结构化数据的自动索引处理数据文件的格式转换，实现不同系统数据的访问，通过编排转换和提取序列描述文本图像，构建注释语料库[36]。

计算档案学档案分析方向主要探讨传统与新兴档案材料的分析问题，包括文本挖掘、数据挖掘、情感分析、网络分析。例如，在文本与数据挖掘方面，T. Blanke 使用“远读”的词频自动化分析和主题建模识别近 80 年英国政府白皮书用语变化特征，并开展档案文本的时代分类和政治模式的演变分析[37]；马里兰州档案馆和马里兰大学数字管护创新中心采用众包形式开展奴隶制遗产项目，以计算方法和开放源码工具将 30 多个档案系列中多类型、分散的文件编码集合，并以可视化工具分析超过 42 万条奴役档案数据间的关系，来反映马里兰州的奴隶制和非裔美国人的真实情况[10]；利默里克大学“埋葬数据”项目将人口普查报告的文本内容转化为细粒度数据，探索 1864－1922 年爱尔兰的历史，并使用机器学习算法来描绘潜在社会结构模式[38]。在情感分析方面，加州大学采用三步社交媒体相似性映射方法自动识别分析已存档的 Twitter 记录，计算与测试集合的情感相似度以筛查新冠肺炎疫情防控期间的各类情绪趋势[39]，如利用机器学习和数据分析揭示和证实新冠肺炎仇恨言论推特档案馆(COVID-19 Hate Speech Twitter Archive，CHSTA)内档案的情感趋势，为危机应对或公共政策的制定提供数据[40]。此外，面对技术应用伴生的伦理难题，计算档案学重视解决数据安全、个人隐私等方面的问题。如荷兰艺术与人文实验室创建 Jupyter Notebooks 归档工具以提供元数据存档和可视化服务，结合法学理论锚定网络环境中档案信息安全和个人隐私保护问题[41]；美国国会图书馆国家录音保存委员会在建立国家广播记录数据库时以政治学视角预设不同主体政治代表性的平衡问题和数据道德操守问题[42]。

计算档案学档案化处理方向主要探讨新兴文件或数据(集)的档案化处理问题，包括识别、元数据生成、完整性检查、规范化、区块链、匿名化等。例如，通过数据集的运算对欧洲文化遗产数字平台的元数据方案完整性进行测试[43]；通过对基因组学数据管理的测试应用证明数据集标识符的分配会提升数据集可用性[44]；通过区块链创新应用于土地交易数据、健康档案和加密货币数据的保存模式归纳了镜像系统、数字记录和

记号代码等维护文件真实性与安全性的保存策略[14]等。另外，T. Miksa 等提出通过对动态的数据信息流与数据模型的充分认识，设计具备可行性的机器可操作数据管理方案，确定支持数据管理任务自动化的必要服务和基础设施组件[45]；H. Hamouda 等结合档案鉴定理论与工程方法开发了 6 种独特的视频测试方式，识别视觉组件、音频组件和元数据组件 3 个关键组件，核查视频内部一致性与外部一致性[46]。

### 2.3　计算档案学的发展仍处于初步探索阶段

自 2016 年以来，国内外多个学科领域学者介入，对计算档案学的理论与实践进行了持续探索，计算档案学内涵逐渐清晰，核心领域正逐步被识别，但不可否认，计算档案学的发展仍处于初步探索阶段。通过对 75 篇外文文献所用研究方法的分类统计（见表 1）发现，当前国外计算档案学研究方法使用具有明显的单一化倾向，非实证研究文献占比高达 73.4%。且非实证研究中案例类文献和介绍类文献占比较大，说明了计算档案学作为一个新兴学科领域尚未形成较为成熟的研究框架和理论体系，其发展初期学界重点关注的是对此领域相关基础问题以及相关实践案例的介绍与分析。

表 1　计算档案学外文文献研究方法分类统计

| 一级类目 | 二级类目 | 解释 | 文献数/篇 | 占比/% |
|---|---|---|---|---|
| 非实证研究 | 介绍类文献 | 介绍计算档案学相关概念、研究和实践情况 | 16 | 21.3 |
| | 观点类文献 | 表达作者的观点、辨析或理解 | 10 | 13.3 |
| | 案例类文献 | 对某一个案进行全面分析与揭示 | 29 | 38.7 |
| 实证研究 | 模型类文献 | 收集数据验证和修正某一理论模型 | 1 | 1.3 |
| | 实验类文献 | 设计实验来测试或验证某些技术方法 | 19 | 25.3 |

笔者进一步对 75 篇外文文献所属研究主题进行分类统计（见表 2），发现当前国外计算档案学研究主题中，基础研究与应用研究呈现出较大的数量差距，基础研究文献仅占 24%，而应用研究文献占 76%，表明计算档案学领域的探索具有很强的“应用性”，致力于解决档案实践当中的计算技术应用问题。具体而言，在应用研究当中，关于档案处理的研究约占一半，其余为档案化处理和档案分析主题；在基础研究当中，对基础理论的研究较多，计算档案学教育问题涉及较少。当前计算档案学领域的很多基本问题尚待解决，例如，计算档案学的研究对象、研究任务、研究范围等缺乏足够探讨。同时由于研究范畴未定，计算档案学与数据科学、数字人文、数字管护等学科或领域的关系与边界也难以明确和区分。计算档案学的研究框架、技术体系、实践路径等相关问题也尚未明晰。

表 2　计算档案学外文文献研究主题分类统计

| 一级类目 | 二级类目 | 类目说明 | 文献数/篇 | 占比/% |
|---|---|---|---|---|
| 基础研究 | 基础理论 | 阐释计算档案学相关概念、特征、学科范围与边界等 | 13 | 17.3 |
| | 教育研究 | 分析计算思维与档案思维相关教育、培训与课程设置问题 | 5 | 6.7 |
| 应用研究 | 档案处理 | 探讨档案材料的处理问题，如电子发现、个人信息识别、鉴定、整理和描述、访问，数字管护，语义、本体、关联数据。主题建模、自然语言处理、机器学习等 | 38 | 50.7 |
| | 档案分析 | 讨论传统与新兴档案材料的分析问题，包括文本挖掘、数据挖掘、情感分析、网络分析 | 7 | 9.3 |
| | 档案化处理 | 讨论新兴文件或数据(集)的档案化处理问题，包括识别、元数据生成、完整性检查、规范化、区块链、匿名化等 | 12 | 16.0 |

当然，如果从学科的视角来审视计算档案学，笔者认为计算档案学尽管有发展成为一门学科的潜力，但目前而言，其尚不具备成为一门学科的条件。当前，国际上虽有专业学术会议来探讨计算档案学问题，美国马里兰大学、加拿大英属哥伦比亚大学和英国伦敦国王学院依托相关研究机构和计算基础设施也零星形成了一些学术团队和合作网络，且他们在计算档案学相关课程建设和人才培养方面进行了大胆的尝试，但当前计算档案学研究成果绝大多数为会议论文，仅有少量期刊论文，专门的学术著作缺乏，也没有创设专门的学术期刊、学术团体和研究机构，计算档案学的学位教育更是空白。计算档案学的可持续发展还面临巨大的挑战。

## 3　中国计算档案学发展的 SWOT 分析

2016 年以来，在多个国家和机构的持续关注以及

核心研究团队的推进下，国外计算档案学蓬勃发展，但在发展初期也面临诸多困境，如学科内涵不清晰、研究范围与边界不确定等。在中国，计算档案学的相关实践与学术研究主要分散在档案数据化、档案数据治理、智慧档案与智慧档案馆建设、区块链与文档管理等方面。中国人民大学信息资源管理学院、上海大学图书情报档案系、四川大学公共管理学院的部分学者在积极追踪国外计算档案学研究动态，但国内计算档案学研究偏向基础理论，应用研究和实践探索明显滞后。当前，国内没有建立起相关的计算档案学研究中心或实验室，也没有形成跨学科的合作研究团队，更没有任何关于计算档案学的国家级科研项目立项，国内学界和业界对于计算档案学能否适应中国学科建设与学术研究环境、如何融入实践发展等问题尚缺明确答案。为此，笔者试图通过 SWOT 分析，参考国外推进计算档案学发展的实践经验，结合国内政策背景、学科建设与实践需求等审视计算档案学在中国发展的内外部环境，以探索计算档案学在中国的发展策略，如表 3 所示：

表 3　中国计算档案学发展的 SWOT 矩阵

| 内外部环境 | 优势(Strengths)<br>• 已确立解决实践问题的两条思路<br>• 已形成初具规模的学科研究方向 | 劣势(Weaknesses)<br>• 计算档案学资源投入还严重不足<br>• 计算档案学研究力量还极为薄弱 |
|---|---|---|
| 机会(Opportunities) | S-O 策略 | W-O 策略 |
| • 符合高等教育与人才培养趋势<br>• 符合档案信息化发展战略要求 | • 依托交叉学科的建设机遇，打造超学科研究平台<br>• 瞄准领域信息化战略需求，形成规模化研究方向 | • 抓住复合型人才培养契机，整合多方资源共建教学平台<br>• 围绕实践领域的核心问题，探索可操作的技术解决方案 |
| 威胁(Threats) | S-T 策略 | W-T 策略 |
| • 受到相关新兴交叉学科的冲击<br>• 受到数据安全风险的较大影响 | • 厘清与相关学科的边界，突出计算档案学的特色<br>• 发挥超学科的研究优势，规避数据安全隐私风险 | • 加强制度设计与技术攻坚，做好档案安全风险评估与管控<br>• 加大基础研究，明晰计算档案学的理论、方法和技术体系 |

## 3.1　中国计算档案学发展的外部环境分析

### 3.1.1　机会：符合高等教育与人才培养趋势

2019 年 5 月，教育部等 13 个部门联合正式启动"六卓越一拔尖"计划 2.0，新文科建设作为此项计划的重要组成部分，坚持以问题为导向，回应社会需求，强调打破学科壁垒，文理相融，进行新专业或新方向、新模式等方面的探索与实践[47]，这不仅反映了当前中国学科发展与高等教育人才培养的新趋势，同时进一步呼唤数字孪生世界下新研究方法和工具的出现[48]。新文科建设下，呼吁将学科问题与数字技术深度融合，实现思辨与计算相结合，提升人文社会科学数据资源的智慧化层次。而计算档案学正是以数据驱动的档案实践与档案研究问题为导向，回应数智时代背景下档案与历史、文化、社会、科学等各方面存在的社会需求，顺应了新文科及文科实验室建设的趋势。这种趋势将促使相应平台和工具的开发与建设，如数据采集处理、数据长期保存、数据可视化等通用性文科实验平台，语义理解、细粒度知识抽取等针对性系统工具等，这些平台和工具将为计算档案学的发展奠定坚实基础。此外，集成实验平台的研发将为计算方法和资源应用于大规模的文件或档案的处理、分析、存储、长期保存和获取工作提供经验，有利于计算思维与档案思维进行融合塑造一个全新的超学科领域。

### 3.1.2　机会：符合档案信息化发展战略要求

2020 年新修订的《中华人民共和国档案法》增加"档案信息化建设"专章，明确了档案信息化建设的总体原则与工作重点，突出了档案信息化建设新要求。在"十四五"期间，中国档案信息化战略将进一步围绕数字档案资源体系、应用系统和利用体系、基础设施与安全体系 3 个方面的任务进行规划设计，且朝着数据化、网络化、自动化、智能化等方向发展。计算档案学以及相应的计算档案实验室的建设正好符合档案信息化战略的要求：一方面，计算档案学致力于提高数据处理效率、生产力和精确度，通过计算方法实现档案数据结构化处理、数据关联等，将为档案数据资源开发、治理、共享和应用提供强力支撑，助推智慧档案应用平台的建设；另一方面，计算档案学致力于解决政府、企业、科研和网络空间等行业或领域新兴数字文件或数据资源的档案化治理、长期保存与维护，为数字文件单轨运行和单套保存、数据连续性保障提供方法指导和系统化、自动化应用解决方案。

### 3.1.3　威胁：受到相关新兴交叉学科的冲击

近些年，数据科学和数字人文等交叉学科在中国发展迅速，研究机构更如雨后春笋般涌现。据教育部统计，截至 2021 年 6 月 30 日，北京大学、清华大学、中国科学技术大学、武汉大学等共 12 所高校设立了数据科学这一交叉学科，涉及的一级学科包括计算机科学与技术、软件工程、管理科学与工程、图书情报与档案管理、数学、统计学、信息与通信工程等。同时，复旦大学、华东师范大学、云南大学、中国人民大学等高校也

在计算机科学与技术、软件工程、管理科学与工程、统计学等一级学科下增设数据科学相关二级学科。中国人民大学在图书情报与档案管理一级学科下增设数字人文二级学科,融合信息资源管理学院、历史学院、国学院、艺术学院、法学院、环境学院等师资队伍,探索跨学科培养数字人文新文科人才的创新道路。相比于数据科学与数字人文这两个学科建设的热度,计算档案学还未受到太多关注,加之与数据科学、数字人文等相关学科边界尚未明确,在学科建设的基础、资源和方向的集聚方面刚兴起的计算档案学易被忽视和受到冲击。

#### 3.1.4 威胁:受到数据安全风险的较大影响

当前,国家对档案数据安全问题尤为重视,但目前档案数据安全法规体系以及顶层设计尚不健全,档案数据面临黑客侵袭等技术风险,导致档案数据的安全存在隐患[49]。档案机构对档案数据安全的隐忧,制约了档案数据的规模化开放与开发,限制了档案数据资源的获取途径和处理成效,导致计算档案学研究的数据资源不足、研究前提不具备。当前,中国档案数据开放形式有限,大多以目录数据形式开放,若要大规模地开放档案内容数据需要从其政策、制度、技术、平台、形式、数据治理成熟度与准备度评估等层面进行全方位推进。由于档案人员技术知识的匮乏与档案机构复合型人才储备的不足,大规模历史档案材料的处理、分析与应用开发都依赖于第三方,导致外包过程中档案数据存在诸多安全隐患。而大部分档案机构害怕承担这种风险,不愿意让馆藏资源脱离其保管场所和管控范畴,对于第三方的介入也保有非常谨慎的态度。这些安全隐患的存在以及档案机构对风险的担忧都会在一定程度上阻碍致力于大规模文件或档案材料处理与研究的计算档案学的发展。



### 3.2 中国计算档案学发展的内部环境分析

#### 3.2.1 优势:已确立解决实践问题的两条思路

当前,计算档案学实践思路已逐渐清晰:一方面,致力于解决档案部门数据化、网络化、自动化和智能化转变过程中遇到的问题。例如,大规模档案材料的敏感性审查、隐私和开放鉴定问题;新兴原生数字态和数据态档案材料的价值鉴定与保存决策问题;以历史、社会、科学、文化研究需求等为导向的大规模档案材料的挖掘和研究问题等,涉及将计算方法和资源运用到大规模文件或档案材料的处理、分析、存储、长期保存和获取。另一方面,致力于解决社会各领域新兴数字文件和(大)数据治理过程中遇到的问题。例如,政府数据资源的长期保存问题、科学大数据的档案化保存问题、数据分析与治理活动过程的文件保存问题、数据态遗产的价值鉴定与长期保存选择标准问题、大数据安全治理与个人隐私保护问题等[50],确保电子档案满足来源可靠、程序规范、要素合规等要求,确保数据资源符合连续性、可追溯性、可信性、可靠性、安全性等要求,涉及将档案理论与方法运用到大数据治理以及各部门新兴数字文件的保存当中。

#### 3.2.2 优势:已形成初具规模的学科研究方向

计算档案学这一新兴超学科领域存在巨大潜力,形成一些初具规模的研究方向,并且已有相关实践作为支撑,例如国外学者通过 8 个案例展示不同交叉学科努力解决档案实践环境的变化,提出构建计算档案学 8 个领域下的应用方式[22],围绕档案材料分析、开发新形式档案、提供档案化处理的拓展服务、用于以档案与馆藏为基础的研究以及网络基础设施建设、大数据和档案的理论与实践等方面初步形成了基础理论、教育研究、档案处理、档案分析、档案化处理 5 个理论研究与实践探索的方向。立足国内实践思路,计算档案学将以档案信息化、档案数据化、智慧档案与智慧档案馆、档案知识发现与知识服务、档案数据治理、档案数据基础设施建设、档案数据开放、档案数据保全、区块链与文档管理、人工智能与文档管理、可信数字文件、档案化与数字管护等方向为基础,整合分散于其中的研究内容,并形成新的统一、整体的认识。

#### 3.2.3 劣势:计算档案学资源投入还严重不足

当前是计算档案学在国外兴起的第五年,以美国马里兰大学信息学院及其数字管护创新中心为首的研究机构围绕计算档案学的建设发展投入了一定数量的资源,包括:网络基础设施等计算资源,启动用于开发维护大规模管理文件数据的数字存储库软件 DRASTIC 计划[51];人财物资源,与美国国家档案与文件署、马里兰州档案馆等档案机构进行合作,由合作的档案机构提供可供分析挖掘的档案资源,马里兰大学信息学院及其学生团队、档案机构工作人员等人员在美国博物馆与图书馆服务协会、美国国家科学基金会等基金的资助下合作开展项目。而国内与计算档案学相关的探索才起步。目前,仅有浙江省档案馆、青岛市档案馆等极少数机构在此领域进行了大胆尝试,例如浙江省档案馆提出建设档案数据中心,并以此为契机与阿里云计算有限公司签署智慧档案研究合作框架协议,攻坚档案开放鉴定、档案数据治理等难题。总体上看,中国在计算档案学资金、计算资源等方面的投入还严重不足,不利于计算档案学的建设与发展。

#### 3.2.4 劣势:计算档案学研究力量还极为薄弱

理想状态下的计算档案学研究人员需要兼备档案思维和计算思维,但目前大部分档案学研究人员缺乏计算思维和熟练运用相关计算方法的技能,而计算机科学研究人员缺乏档案思维。当前,国外计算档案学的研究力量主要来自高校研究机构、档案机构,形成了跨界合作模式,实现资源和技术的互补。而目前国内计算档案学研究力量主要来自高校研究机构,缺少跨学科研究团队。一方面,档案学界并未明晰统一计算档案学内涵与边界,跨学科的交流与合作存在极大阻碍。不同学科对计算档案学的重视程度不同,其参与定位和角色动机不够明朗。另一方面,档案学界并未提出具有普适性和系统性的复合型档案人才培养体系。面对兼备档案思维和计算思维的要求,国内需要以跨界合作的观念组建跨学科团队共同开展计算档案学的研究和实践,围绕计算档案学应用计算方法和资源处理分析大规模文件/档案以提高效率和精确度的核心目的,探索契合计算思维方法与档案思维方法的研究方法。

## 4 中国计算档案学发展的策略研究

### 4.1 优势——机会(SO)策略

首先,要依托交叉学科的建设机遇,打造超学科研究平台。培养计算机与档案复合型人才,不仅需要档案学研究人员的参与,还需要与其他学科加强跨领域合作,与计算机科学、信息科学等学科共同组建研究团队,抓住国家鼓励学科融合建立新学科的新文科建设机遇,利用政策优势,积极探索建设超学科的方法创新,建立跨学院、跨学科的超学科研究平台。其次,要瞄准领域信息化战略需求,形成规模化研究方向。国外学者提出的计算档案学领域研究主题与方向涉及面甚广,如大数据、图形分析、数字人文,网络基础设施建设等。而面向国内,需要考察中国政府信息化、档案信息化、文化信息化等领域信息化战略需求与现实问题,对接计算档案学的理论、思维与方法输出,重点关注档案数据开发与利用、数据连续性保障、档案数据管护与治理等问题,以实践需求集聚与壮大具有国内特色的计算档案学研究方向。

### 4.2 优势——威胁(ST)策略

首先,要厘清与相关学科的边界,突出计算档案学的特色。计算档案学作为“计算＋”学科阵列当中的新成员,是基于档案学、信息科学、计算机科学等学科要素重组而创造出新知识的超学科领域。由于当前计算档案学学科框架还未完全形成,尚存在与数字人文、数据科学等相关学科重合的领域,与其差异与边界也尚未形成较为明确的认知。要弄清计算档案学与相关学科的边界,首先要明确计算档案学的学科体系和研究框架,在此基础上可以抓住学科内涵探究学科发展范围与领域,探寻未来发展方向,同时可以找准与数字人文、数据科学等相关学科的明确区别,形成与计算社会科学、计算语言学、计算情报学等相区别的独立身份和特征,突出计算档案学的特色。其次,要发挥超学科的研究优势,规避数据安全隐私风险与技术伦理风险。数智时代的到来虽极大推动了社会的变革,但技术应用也加剧数据安全风险,同时伴生道德伦理问题,为国家和领域数据安全保障带来严峻挑战。无论是政府开放数据、科学数据、研究数据还是档案部门管控的档案数据,要实现数据资源的开放、开发、交易、利用、共享等,前提是解决好数据安全、保密和隐私等问题,尤其是不能危害总体国家安全。计算档案学是计算理论方法与档案理论方法的有机融合,对于解决数据安全和隐私问题有先天优势。因此,在计算档案学领域的发展当中,要发挥出这种超学科的研究优势,为规避和防控数据安全和隐私风险提供理论、方法和技术支撑。

### 4.3 劣势——机会(WO)策略

首先,要抓住复合型人才培养契机,整合多方资源共建教学平台。21 世纪以来,在信息技术的驱动下,既懂技术又懂管理的复合型人才需求急剧攀升,政府信息化、档案信息化等领域尤甚,但一直以来,中国档案学或公共管理等学科的人才培养当中并未有效解决复合型人才培养的问题。计算档案学的出现为弥合兼具计算思维和档案思维的复合型人才培养缺口带来了契机。应先进行跨学科教学平台搭建,集中多学科师资力量,讨论制定计算档案学课程体系,将不同学科的知识融入教学课程中[22],全面改造档案学传统核心专业课程内容,开发培养计算思维的档案技术类应用课程等。同时,整合学科文献资源、技术设备资源、校内外合作资源等多种资源建立学科实践平台,借鉴国外 iSchool 项目式培养模式,建设产学研创新基地,让学生参与计算档案学项目的实践,从实践中掌握计算档案学相关知识和方法。其次,要围绕实践领域的核心问题,探索可操作的技术解决方案。将实践需求转化为学科发展的问题导向,围绕领域信息化发展战略需求和领域数据管理的核心问题,依托计算档案学探索可操作的技术解决方案。例如,围绕区块链技术的应用探索可信数字档案建设的解决方案,围绕电子档案的四性检测问题探索一体化的四性检测工具,围绕人工智能技术的应用探索开放档案智能鉴定的解决方

案等。

### 4.4 劣势——威胁(WT)策略

首先,要加强制度设计与技术攻坚,做好档案安全风险评估与管控。自 21 世纪以来,随着经济社会的快速发展,档案工作所处的内外部环境日趋复杂,危及档案安全的传统风险与非传统风险日益增多,确保档案安全成为中国档案事业发展的重要内容。当前,在各领域开放运动背景下,从档案开放走向档案数据开放,需要依据新修订档案法的规定,做好自上而下的开放政策与制度设计以及自下而上的数据开放技术攻坚,重点解决档案数据开放中的开放与保密的矛盾、处理好数据开放与隐私保护的关系;以项目为依托,以自然语言处理、机器学习等技术为切入点,针对文本、图像、音频、视频的不同类型以及电子邮件、电子公文、网页文件、社交媒体文件等不同形式的处理对象,加强开放和安全鉴定的技术攻关;此外,要做好档案数据安全风险评估,识别潜在的、可能发生的档案数据安全风险要素,防范档案形成机构及其工作人员的风险性行为,及时消除档案安全隐患,以弥补或减少损失。同时,在认同计算思维和技术提升数据处理能力的同时重视数字鸿沟的风险,加强数字包容。面向大规模、多类型的文件与数据处理优势在解决不同实际需求的基础上,能注重消弭不同群体间共享资源的距离。其次,要加大基础研究的投入与支持,明晰计算档案学的理论、方法和技术体系。中国自 21 世纪初以来规模化开展的档案数字化工程,为计算档案学奠定了良好的基础,一方面如何在繁琐冗余的低价值密度数据中通过数据方法剥离出有价值的信息,另一方面又要如何避免技术给档案带来的消极影响,如何在新技术环境下理解文件,解决这些问题都需要档案工作者和计算机科学家的深入合作。而目前计算档案学发展时间短暂,学科发展尚未成熟,应加大对其基础研究的投入,明确计算档案学基础理论和方法论,从而克服计算档案学实践应用的潜在问题。

## 5 结语

在数智时代,图情档学科需要有新的社会贡献力,并且亟需在新时期发出自己的声音。数智时代的到来与伴生的变化为图情档带来新的挑战,档案学科和档案职业必须广泛关注并积极回应社会的重大挑战。面对信息化与大数据管理的实践需求,依靠档案学知识已无法准确高效回应,档案学者们已经关注到档案学科必须和其他学科合作共同应对新挑战。计算档案学的发展提供了学科间交流、合作、融合的平台。计算档案学不仅是计算机科学向档案学单向输出方法、技术,而是档案学与其他学科之间多向输出,档案学科应抓住计算档案学发展的契机输出档案学科知识、理论与方法,扩大档案学科影响力。计算档案学在国际上的兴起并非偶然,新的档案形式的不断出现,呼吁多学科领域的共同应对;传统档案工作加速转型,也要求计算思维与方法的介入;复合型档案人才缺口较大,更需要创新档案高等教育方式。当前,经过国内外学者的探索,计算档案学超学科内涵逐渐清晰,核心领域正逐步被识别,但其学科边界、研究框架、技术体系、实践路径等尚未明晰,仍需进一步探讨。本文基于 SWOT 分析法在一定程度上提出了中国计算档案学的发展策略,但由于计算档案学的发展尚处于初步阶段,尤其是中国计算档案学的建设并未在实践层面全面展开,因此本文利用 SWOT 定性分析得到的结论偏于宏观和主观,存在不足,未来的研究还可进一步结合多学科专家咨询或深度访谈等方法了解图情档学界和业界对中国推进计算档案学的态度或建议。

**参考文献:**

[1] CAS WORKSHOP. IEEE big data 2016: CAS #1[EB/OL]. [2021-01-20]. https://ai-collaboratory.net/cas/cas-workshops/ieee-big-data-2016-1st-cas-workshop/.
[2] 傅天珍,郑江平. 计算档案学的兴起、探索与启示[J]. 档案学通讯, 2019(4): 28-33.
[3] 周文泓,代林序,贺谭涛,等. 计算档案学的内涵解析与展望[J]. 档案学研究,2021(1):49-57.
[4] 陶玉芳. 计算档案学的研究现状及未来展望[J]. 浙江档案,2021(2):53-56.
[5] 刘越男,杨建梁,何思源,等. 计算档案学:档案学科的新发展[J]. 图书情报知识,2021,38(3):4-13.
[6] 于英香,刘茜. 论计算档案学的出场逻辑[J]. 档案学通讯,2021(5):22-31.
[7] 赵跃,张佳欣. 计算档案学在中国的发展前景探析——基于中国图情档界的计算档案学认知调查[J]. 档案学通讯,2021(5):32-39.
[8] UMD ISCHOOL. Research centers & labs[EB/OL]. [2021-11-24]. https://www.ischool.umd.edu/research/centers-and-labs.
[9] UMD ISCHOOL. Directory of richard marciano[EB/OL]. [2021-11-24]. https://www.ischool.umd.edu/about/directory/richard-marciano.
[10] COX R, SHAH S, FREDERICK W, et al. A case study in creating transparency in using cultural big data: the legacy of slavery project[C]// 2018 IEEE international conference on big data. Piscataway: IEEE, 2018: 2689-2695.
[11] UNDERWOOD W, MARCIANO R, LIAB S, et al. Computational curation of a digitized record series of WWII Japanese-American internment[C]// 2017 IEEE international conference on big data. Piscataway: IEEE, 2017: 2309-2313.

[12] MARCIANO R, LEE M, UNDERWOOD W, et al. Digital curation of a World War II Japanese-American incarceration camp collection: implications for sociotechnical archival systems [C]// 2018 IEEE international conference on big data. Piscataway: IEEE, 2018: 1 –4.

[13] BATISTA D, WEINGAERTNER T. ArchContract: using smart contracts for disposition[C]// 2019 IEEE international conference on big data. Piscataway: IEEE, 2019: 3060 –3065.

[14] LEMIEUX V. A typology of blockchain recordkeeping solutions and some reflections on their implications for the future of archival preservation[EB/OL]. [2021 – 09 – 30]. https://dcicblog. umd. edu/cas/wpcontent/uploads/sites/13/2017/06/Lemieux. pdf.

[15] PAYNE N. Auto-categorization & future access to digital archives [EB/OL]. [2021 – 04 – 13]. https://dcicblog. umd. edu/cas/wp-content/uploads/sites/13/2017/06/Payne. pdf.

[16] PAYNE N. An intelligent class: the development off a novel context capturing framework for the functional classification of records [EB/OL]. [2021 – 04 – 13]. https://dcicblog. umd. edu/cas/wpcontent/uploads/sites/13/2019/11/Payne. pdf.

[17] KING' S COLLEGE LONDON. About the department of digital humanities[EB/OL]. [2021 –11 –24]. https://www. kcl. ac. uk/ddh/about/about.

[18] BRYANT M. GraphQL for archival metadata: an overview of the EHRI GraphQL API [EB/OL]. [2021 – 04 – 13]. http://dcicblog. umd. edu/cas/wpcontent/uploads/sites/13/2017/06/Bryant. pdf.

[19] SOBIGDATA. European research infrastructure for big data and social mining[EB/OL]. [2021 –11 –24]. http://www. sobigdata. eu/index.

[20] NOAH D. Computational archival science international network to be launched[EB/OL]. [2021 –11 –24]. https://ischool. umd. edu/news/computational-archival-science-international-network-be-launched.

[21] MARCIANO R, LEMIEUX V, HEDGES M, et al. Archival records and training in the age of big data[M]// PERCELL J, SARIN L C, JAEGER P T, et al. Re-envisioning the MLS: perspectives on the future of library and information science education, Bingley: Emerald, 2018: 179 –199.

[22] PAYNE N. Stirring the cauldron: redefining computational archival science (CAS) for the big data domain[C]// 2018 IEEE international conference on big data. Piscataway: IEEE, 2018: 2743 –2752.

[23] LEE M, ZHANG Y, CHEN S, et al. Heuristics for assessing computational archival science (CAS) research: the case of the human face of big data project[C]// 2017 IEEE international conference on big data. Piscataway: IEEE, 2017: 2262 –2270.

[24] STANCIC H. Computational archival science [EB/OL]. [2021 – 09 – 30]. http://bib. irb. hr/datoteka/994072. Stancic_H._Computational_Archival_Science. pdf.

[25] THIBODEAU K. Computational archival practice: towards a theory for archival engineering [EB/OL]. [2021 – 04 – 12]. https//dcicblog. umd. edu/cas/wpcontent/uploads/sites/13/2018/12/3. Thibodeau. pdf.

[26] STANCIC H, RAJH A, JAMIC M. Impact of ICT on archival practice from the 2000s onwards and the necessary changes of archival science curricula[C]// Proceedings of the 40th Jubilee international convention on information and communication technology, Electronics and microelectronics MIPRO 2017. BiljanoviÄ: Petar, 2017:812 –817.

[27] WEINTROP D, BEHESHTI E, HORN M, et al. Defining computational thinking for mathematics and science classrooms[J]. Journal of science education and technology, 2015, 25 (1): 127 –147.

[28] UNDERWOOD W, WEINTROP D, KURTZ M, et al. Introducing computational thinking into archival science education[C]// 2018 IEEE international conference on big data. Piscataway: IEEE, 2018: 2761 –2765.

[29] UNDERWOOD W, MARCIANO R. Computational thinking in archival science research and education[C]// 2019 IEEE international conference on big data. Piscataway: IEEE, 2019: 3146 –3152.

[30] MARCIANO R, JANSEN G, UNDERWOOD W. Developing a framework to enable collaboration in computational archival science education [EB/OL]. [2021 – 04 – 13]. https://www2. archivists. org/sites/all/files/Marciano,%20Richard. pdf.

[31] DUGENIE P, FREIRE N, BROEDER D. Building new knowledge from distributed scientific corpus: HERBADROP & EUROPEANA: two concrete case studies for exploring big archival data [C]// IEEE international conference on big data. Piscataway: IEEE, 2017: 2231 –2239.

[32] SHALLCROSS M. Appraising digital archives with archivematica [EB/OL]. [2021 –04 –13]. https://dcicblog. umd. edu/cas/wp-content/uploads/sites/13/2016/05/9. pdf.

[33] HUTCHINSON T. Protecting privacy in the archives: preliminary explorations of topic modeling for born-digital collections [EB/OL]. [2021 –04 –11]. https://dcicblog. umd. edu/cas/wp-content/uploads/sites/13/2017/06/Hutchinson. pdf.

[34] HUTCHINSON T. Protecting privacy in the archives: supervised machine learning and born-digital records [EB/OL]. [2021 –04 –13]. http://dcicblog. umd. edu/cas/wp-content/uploads/sites/13/2018/12/5. Hutchinson. pdf.

[35] BRYANT M. In-place synchronisation of hierarchical archival descriptions [EB/OL]. [2021 – 04 – 13]. https://dcicblog. umd. edu/cas/wp-content/uploads/sites/13/2018/12/8. Bryant. pdf.

[36] THOMAS W R. Petabytes in practice: working with collections as data at Scale[J]. Data and information management, 2019;3(1): 18 –25.

[37] BLANKE T. Identifying epochs in text archives [EB/OL]. [2021 – 04 – 13]. https://dcicblog. umd. edu/cas/wp-content/uploads/sites/13/2017/06/Blanke. pdf.

[38] O'SHEA E, KHAN R, BREATHNACH C, et al. Towards automatic data cleansing and classification of valid historical data an incremental approach based on MDD[C]// IEEE international conference on big data. Piscataway: IEEE, 2020:1914 –1923.

[39] YIN Z, FAN L, YU H, et al. Using a three-step social media similarity (TSMS) mapping method to analyze controversial speech re-

lating to COVID-19 in twitter collections[C]// IEEE international conference on big data. Piscataway: IEEE, 2020:1949 - 1953.
[40] FAN L, YIN Z, YU H, et al. Using data-driven analytics to enhance archival processing of the COVID-19 hate speech twitter archive (CHSTA)[EB/OL]. [2021 - 04 - 13]. https://www.researchgate.net/publication/349071485_Using_machine_learning_to_predict_concrete's_strength_Learning_from_small_datasets.
[41] WIGHAM. Jupyter notebooks for generous archive interfaces[EB/OL]. [2021 - 04 - 13]. https://dcicblog.umd.edu/cas/wp-content/uploads/sites/13/2018/12/13.Wigham.pdf.
[42] GOODMANN E, Matienzo M A, Vancour S, et al. Building the national radio recordings database: a big data approach to documenting audio heritage[C]//2019 IEEE international conference on big data. Piscataway: IEEE, 2019: 3080 - 3086.
[43] KIRALY P. Measuring completeness as metadata quality metric in europeana[EB/OL]. [2021 - 09 - 13]. https://dcicblog.umd.edu/cas/wp-content/uploads/sites/13/2018/12/7.Kiraly.pdf.
[44] XU W, HHUANG R, ESTEVA M, et al. Content-based comparison for collections identification[C]// IEEE international conference on big data. Piscataway: IEEE, 2016: 3283 - 3289.
[45] MISKA T, CARDOSO J, BORBINHA J. Framing the scope of the common data model for machine-actionable data management plans[C]// 2018 IEEE international conference on big data. IEEE, 2018: 2733 - 2742.
[46] HAMOUDA H, BUSHEY J, LEMIEUX V, et al. Extending the scope of computational archival science: a case study on leveraging archival and engineering approaches to develop a framework to detect and prevent "fake video"[C]// 2019 IEEE international conference on big data. Piscataway: IEEE, 2019: 3087 - 3097.
[47] 樊丽明."新文科":时代需求与建设重点[J].中国大学教学,2020(5):4 - 8.
[48] 唐衍军,蒋翠珍.跨界融合:新时代新文科人才培养的新进路[J].当代教育科学,2020(2):71 - 74.
[49] 金波,杨鹏.大数据时代档案数据安全治理策略探析[J].情报科学,2020,38(9):30 - 35.
[50] 赵跃,孙晶琼,段先娥.档案化:档案科学介入数据资源管理的理性思考[J].档案学研究,2020(5): 83 - 91.
[51] CAS WORKSHOP. Computational archival science [EB/OL]. [2021 - 10 - 21]. https://dcicblog.umd.edu/cas/4-cas-cyberinfrastructure/.

**作者贡献说明:**
赵跃:提出研究问题、思路,撰写、修改论文;
马晓玥:撰写、修改论文;
张佳欣:分析文献,撰写论文。

## SWOT Analysis and Strategy Research on the Development of Computational Archival Science in China

Zhao Yue[1] Ma Xiaoyue[1] Zhang Jiaxin[2]
[1] School of Public Administration, Sichuan University, Chengdu 610065
[2] School of Information Resource Management, Renmin University of China, Beijing 100872

Abstract: [**Purpose/significance**] This paper reviews the current situation of the development of computational archival science (CAS), explores the development strategies of CAS in China, and provides a reference for the development of CAS in China under the background of the construction of new liberal arts. [**Method/process**] Based on literature research, this paper sorted out the current development situation of CAS, and analyzed the internal and external environmental factors such as opportunities, threats, advantages, and disadvantages of the development of CAS in China by SWOT analysis method, and formed different development strategies through the cross matching of internal and external factors. [**Result/conclusion**] It is found that the development of CAS has received sustained attention internationally and has initially established the direction of theoretical research and practical exploration around basic theory, educational research, archives processing, archives analysis, and archival processing, but in general, the development of CAS is still in the preliminary exploration stage. This paper proposes that, in order to promote CAS in China, we should pay attention to the following aspects: relying on the opportunities of interdisciplinary construction to build a platform for transdisciplinary research; aiming at the strategic needs of field informatization, and forming a large-scale research direction; clarifying the boundaries with related disciplines, and highlighting the characteristics of CAS; taking advantages of transdisciplinary research to avoid data security privacy risks; seizing the opportunities of cultivating interdisciplinary talents, and integrating multiple resources to build a teaching practice platform; exploring operable technical solutions around the core problems in the field of practice; strengthening institutional design and technical breakthroughs, and doing well in archives security risk assessment and control; and increasing basic research, and clarifying the theories, methods and technical systems of CAS.

Keywords: computational archival science new liberal arts data transition archival education